

An Online Convex Optimization Approach to Blackwell's Approachability

Nahum Shimkin

Department of Electrical Engineering
Technion – Israel Institute of Technology
Haifa 32000, ISRAEL
shimkin@ee.technion.ac.il

March 3, 2015

Abstract

The notion of approachability in repeated games with vector payoffs was introduced by Blackwell in the 1950s, along with geometric conditions for approachability and corresponding strategies that rely on computing *steering directions* as projections from the current average payoff vector to the (convex) target set. Recently, Abernethy, Batlett and Hazan (2011) proposed a class of approachability algorithms that rely on the no-regret properties of Online Linear Programming for computing a suitable sequence of steering directions. This is first carried out for target sets that are convex cones, and then generalized to any convex set by embedding it in a higher-dimensional convex cone. In this paper we present a more direct formulation that relies on the support function of the set, along with suitable Online Convex Optimization algorithms, which leads to a general class of approachability algorithms. We further show that Blackwell's original algorithm and its convergence follow as a special case.

1 Introduction

Both Blackwell's theory of approachability and the no-regret framework of online learning address a repeated decision problem in the presence of on an arbitrary (namely, unpredictable) adversary. The concept of approachability, introduced in [4], addresses a fundamental feasibility issue in for repeated matrix games with vector-valued payoffs. Referring to one player as the *agent* and to the other as *Nature*, a set S in the payoff space is *approachable* by the agent if he can ensure that the average payoff vector converges (with probability 1) to S ,

irrespectively of Nature’s strategy. Blackwell provided in his paper geometric conditions for approachability, which are both necessary and sufficient for *convex* target sets S , and a corresponding approachability strategy for the agent. An extensive recent survey of approachability and its implications can be found in [12], and a textbook exposition is available in [11].

Concurrently, Hannan [7] introduced the concept of no-regret play for repeated matrix games. The *regret* of the agent is the shortfall of the cumulative payoff that was actually obtained relative to the one that could have been obtained with the best (fixed) action in hindsight, given Nature’s observed actions. A no-regret strategy, or algorithm, should ensure that the regret grows sub-linearly in time. The no-regret criterion has been widely adopted during the last two decades by the machine learning community as a standard measure for the performance of online learning algorithms, and its scope has been greatly extended. Of specific relevance here is the Online Convex Optimization (OCO) framework, where Nature’s discrete action is replaced by the choice of a convex function at each stage, and the agent’s decision is a point in a convex set. The textbook [6] offers a broad overview of regret and online learning. Recent surveys of OCO algorithms may be found in [15, 9].

It is well known that no-regret strategies for repeated games can be obtained as a special case of the approachability problem. This was already observed in [3]; an alternative formulation that leads to more explicit strategies was proposed in [8]. More recently, it was shown in [1] that any no-regret algorithm for the online *linear* optimization problem can be used as a basis for an approachability strategy for convex target sets. The online algorithm is used here compute a sequence of *steering vectors*, that replace the projection directions used in Blackwell’s original algorithm.

The scheme suggested in [1] first considers target sets S that are convex cones. The generalization to any convex set is carried out by embedding the original target set in a convex cone in a higher dimensional payoff space. The present paper proposes a more direct scheme that avoids the above-mentioned embedding. This is done by invoking the *support function* of the target set, along with well-known relations between this function and the Euclidean distance to the set. As the support function is convex, the full arsenal of OCO algorithms may be applied to provide the required sequence of steering vectors.

A natural question concerns the relation between Blackwell’s original algorithm and the present framework. We first observe that Blackwell’s algorithm is recovered when the standard Follow the Leader (FTL) algorithm is used for the OCO part. Establishing the (known) convergence of this algorithm via the proposed OCO framework is a bit more intricate. First, when the target set has a smooth boundary, we show that FTL guarantees logarithmic rate, which “fast” approachability at a rate of $O(\frac{\log T}{T})$. To address the general case, we further observe that Blackwell’s algorithm is still obtained when a regularized version of FTL is employed,

from which the standard $O(t^{-1/2})$ convergence rate may be deduced.

The paper proceeds as follows. In Section 2 we recall the relevant background on Blackwell’s approachability and Online Convex Optimization. Section 3 presents the proposed scheme, in the form of a meta-algorithm that relies on a generic OCO algorithm, discusses the relation to the scheme of [1], and demonstrates a specific algorithm that is obtained by using Generalized Gradient Descent for the OCO algorithm. In Section 4 we outline the relations with Blackwell’s original algorithm, and provide some concluding remarks.

Notation: The standard inner product in \mathbb{R}^d is denoted by $\langle \cdot, \cdot \rangle$, $\| \cdot \|$ is the Euclidean norm, and $d(r, S) = \inf_{s \in S} \|r - s\|$ denotes the corresponding point-to-set distance. Further, $B_2 = \{w \in \mathbb{R}^d : \|w\| \leq 1\}$ denotes the Euclidean unit ball, $\Delta(I)$ is the set of probability distributions over a finite set I , $\text{diam}(S) = \sup_{s, s' \in S} \|s - s'\|$ is the diameter of the set S , and $\|\mathcal{R} - S\| = \sup_{r \in \mathcal{R}, s \in S} \|r - s\|$ denotes the maximal distance between points in the sets \mathcal{R} and S .

2 Model and Background

We start with a brief review of Blackwell’s approachability and of Online Convex Programming, focusing on those aspects that are relevant to this paper.

2.1 Approachability

Consider a repeated game with *vector-valued* rewards that is played by two players, the *agent* and *Nature*. Let I and J denote the finite action sets of these players, with corresponding mixed actions $x = (x(1), \dots, x(|I|)) \in \Delta(I)$ and $y = (y(1), \dots, y(|J|)) \in \Delta(J)$. Let $r : I \times J \rightarrow \mathbb{R}^d$ be the vector-valued reward function of the single-stage game, which is extended to mixed action as usual through the bilinear function

$$r(x, y) = \sum_{i, j} x(i)y(j)r(i, j).$$

Similarly, we denote

$$r(x, j) = \sum_i x(i)r(i, j).$$

The game is repeated in stages $t = 1, 2, \dots$, where at stage t actions i_t and j_t are chosen by the players, and the reward vector $r(i_t, j_t)$ is obtained. A pure strategy for the agent is a mapping from each possible history $(i_1, j_1, \dots, i_{t-1}, j_{t-1})$ to an action i_t , and a mixed strategy is a probability distribution over the pure strategies. Nature’s strategies may be similarly defined.

As usual, we restrict attention to so-called behavior strategies of the agent, where the action i_t is drawn randomly according to a mixed action x_t , using independent draws across stages. Furthermore, to simplify the presentation, we shall state our results and algorithms in terms of the *smoothed* reward vectors $r(x_t, j_t)$, where the reward $r(i, t, j_t)$ is averaged over the mixed action x_t . This will allow us to state the results in simpler sample-path terms, rather than probabilistic ones; we further discuss this formulation below after Theorem 1.

Let

$$\bar{r}_T = \frac{1}{T} \sum_{t=1}^T r(x_t, j_t)$$

denote the T -stage average reward vector.

Definition 2.1 (Approachability) *A closed set $S \subset \mathbb{R}^d$ is **approachable** if there exists a strategy for the agent and a sequence $\epsilon(T) \rightarrow 0$ such that*

$$\lim_{T \rightarrow \infty} d(\bar{r}_T, S) \leq \epsilon(T) \quad (1)$$

holds (w.p. 1) for any strategy of Nature. A strategy of the agent that satisfies this property is an approachability strategy for S .

Theorem 1 (Blackwell, 1956) *A closed and convex set $S \subset \mathbb{R}^d$ is approachable if and only if either one of the following equivalent conditions holds:*

(i) *For each unit vector $u \in \mathbb{R}^d$, there exists a mixed action $x = x_S(u) \in \Delta(I)$ such that*

$$\langle u, r(x, j) \rangle \leq \sup_{s \in S} \langle u, s \rangle, \quad \text{for all } j \in J. \quad (2)$$

(ii) *For each $y \in \Delta(J)$ there exists $x \in \Delta(I)$ such that $r(x, y) \in S$.*

If S is approachable, then the following strategy is an approachability strategy for S :

For $v \notin S$, let $u_S(v)$ be the unit vector that points to v from $\text{Proj}_S(v)$, the closet point to v in S . Then, for $t \geq 1$, if $\bar{r}_t \notin S$, choose $x_{t+1} = x_S(u_S(\bar{r}_t))$; otherwise, choose an arbitrarily action.

The approachability strategy introduced by Blackwell has been generalized in [8], that essentially allow different norms to be used for the projection unto S . Several recent papers have proposed approachability algorithms that depend on Blackwell's dual condition (condition (ii) in the above Theorem) and avoid the projection step altogether (see [2] and references therein). The current paper again proposes a generalization of Blackwell's strategy, but from a different viewpoint.

Let us elaborate on the use of the smoothed rewards $r(x_t, j_t)$. This offers several useful benefits:

1. As noted, we obtain sample-path bounds rather than probabilistic ones.
2. We can state results that hold for any sequence (j_t) , rather than any (mixed) strategy of Nature. This is closer to the spirit of Online Algorithms, where the notion of a randomized choice by Nature may not be meaningful.
3. As is well known, the difference $\sum_{t=1}^T r(x_t, j_t) - \sum_{t=1}^T r(i_t, j_t)$ is a Martingale difference sequence, hence of order \sqrt{T} . Thus, the difference in the means is of order $\frac{1}{\sqrt{T}}$, and convergence results derived for the smoothed mean are valid for the non-smoothed one up to that order.

We note that the results in [1] are developed for the rewards $r(x_t, y_t)$, with the mean taken over y_t as well, and the agent is allowed to observe Nature's mixed action y_t (or at least the mean reward $r(x_t, y_t)$). We avoid making that extra step and assume that the agent only observes Nature's pure actions $\{j_t\}$.

As the pure actions i_t of the agent do not affect the rewards $r_t = r(x_t, j_t)$, we may suppress them in the following discussion and focus on the mixed actions x_t . In particular, we restrict attention to strategies of the agent that assign a mixed action x_t to each sequence (j_1, \dots, j_{t-1}) of Nature's actions. (Note that there is no need to include the past mixed actions x_1, \dots, x_{t-1} in the history sequence, since they may be computed recursively; in practice, however, we will express x_t as a function of past the reward vector sequence $(r(x_k, j_k))_{k < t}$.) Since there is no randomization involved, it may be seen that Definition 2.1 is equivalent to the requirement that the bound (1) holds (deterministically) for any sequence (j_1, j_2, \dots) of Nature's actions.

2.2 Online Convex Optimization (OCO)

OCO extends the framework of no-regret learning to function minimization. Let W be a convex and compact set in \mathbb{R}^d , and let \mathcal{F} be a set of convex and uniformly bounded functions $f : W \rightarrow \mathbb{R}$. Consider a sequential decision problem, where at each stage $t \geq 1$ the agent chooses a point $w_t \in W$, and then observes a function $f_t \in F$. An *Algorithm* for the agent is a rule for choosing w_t , $t \geq 1$, based on the history $\{f_k, w_k\}_{k \leq t-1}$. The regret of an algorithm \mathcal{A} is defined as

$$\text{Regret}_T(\mathcal{A}) = \sup_{f_1, \dots, f_T \in \mathcal{F}} \left\{ \sum_{t=1}^T f_t(w_t) - \min_{w \in W} \sum_{t=1}^T f_t(w) \right\}, \quad (3)$$

where the supremum is taken over all possible functions $f_t \in \mathcal{F}$. An effective algorithm should guarantee a small regret, and in particular one that grows sub-linearly in T .

The OCO problem was introduced in this generality in [16], along with the following Online

Gradient Descent algorithm:

$$w_{t+1} = \text{Proj}_W(w_t - \eta_t g_t). \quad (4)$$

Here g_t is an arbitrary element of $\partial f_t(w_t)$, the subdifferential of f_t at w_t , (η_t) is a diminishing gain sequence, and Proj_W denotes the Euclidean projection onto the convex set W . To state a regret bound for this algorithm, let $\text{diam}(W)$ denote the diameter of W , and suppose that all subgradients of the functions f_t are uniformly bounded in norm by a constant G .

Proposition 2 (Zinkevich, 2003) *For the Online Gradient Descent algorithm in (4) with gain sequence $\eta_t = \frac{\eta}{\sqrt{t}}$, $\eta > 0$, the regret is upper bounded by*

$$\text{Regret}_T(\text{OGD}) \leq \left(\frac{\text{diam}(W)^2}{\eta} + 2\eta G^2 \right) \sqrt{T}. \quad (5)$$

Several classes of OCO algorithms are now known, as surveyed in [6, 15, 9]. Of particular relevance here is the Regularized Follow the Leader (RTFL) algorithm, specified by

$$w_{t+1} = \underset{w \in W}{\text{argmin}} \left(\sum_{k=1}^t f_k(w) + R_t(w) \right), \quad (6)$$

where $R_t(w)$, $t \geq 1$ is a sequence of regularization functions. With $R_t \equiv 0$, the algorithm reduces to the basic Follow the Leader (FTL) algorithm, which does not generally lead to sub-linear regret, unless additional requirements such as strong convexity are imposed on the functions f_t (we will revisit the convergence of FTL in Section 4). For RFTL, we will require the following standard convergence result. Recall that a function $R(w)$ over a convex set W is called ρ -strongly convex if $R(w) - \frac{\rho}{2}\|w\|^2$ is convex there.

Proposition 3 *Suppose that each function f_t is Lipschitz-continuous over W , with Lipschitz coefficient L_f . Let $R_t(w) = \rho_t R(w)$, where $0 < \rho_t < \rho_{t+1}$, and the function $R : W \rightarrow [0, R_{\max}]$ is Lipschitz continuous with coefficient L_R , and is 1-strongly convex. Then,*

$$\text{Regret}_T(\text{RFTL}) \leq 2L_f \sum_{t=1}^T \frac{L_f + (\rho_t - \rho_{t-1})L_R}{\rho_t + \rho_{t-1}} + \rho_T R_{\max}. \quad (7)$$

The last bound can be established along the lines of Theorem 2.11 in [15], which considers the case of fixed regularization parameters, $\rho_t \equiv \rho_0$. The proof is outlined in the Appendix.

3 OCO-Based Approachability

This section presents the proposed OCO-based approachability algorithm. We start by introducing the support function and some of its properties, and expressing Blackwell's separation

condition in terms of this function. We continue to present the proposed meta-algorithm that employs a generic OCO algorithm, and then provide as an example the specific algorithm that is obtained when Online Gradient Descent is used as the OCO algorithm.

3.1 The Support Function

Let set $S \subset \mathbb{R}^d$ be a closed and convex set. The *support function* $h_S : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ of S is defined as

$$h_S(w) \triangleq \sup_{s \in S} \langle w, s \rangle, \quad w \in \mathbb{R}^d.$$

It is evident that h_S is a convex function (as a pointwise supremum over linear functions), and is positive homogeneous: $h_S(aw) = ah_S(w)$ for $a \geq 0$. Furthermore, the Euclidean distance from a point r to S can be expressed as

$$d(r, S) = \max_{w \in B_2} \{ \langle w, r \rangle - h_S(w) \}, \quad (8)$$

where B_2 is the closed Euclidean unit ball (see, e.g., [5], Section 8.1.3; this equality may be readily verified using the minimax theorem). It follows that

$$\operatorname{argmax}_{w \in B_2} \{ \langle w, r \rangle - h_S(w) \} = \begin{cases} 0 & : r \in S \\ u_S(r) & : r \notin S \end{cases} \quad (9)$$

with $u_S(r)$ as defined in Theorem 1, namely the unit vector pointing to r from $\operatorname{Proj}_S(r)$.

Blackwell's separation condition in (2) can now be written in terms of the support function, as

$$\langle w, r(x, j) \rangle \leq \sup_{s \in S} \langle w, s \rangle \equiv h_S(w).$$

We thus obtain the following Corollary to Theorem 1.

Corollary 4 *A closed and convex set S is approachable if and only if for every vector $w \in B_2$ there exists $x \in \Delta(I)$ so that*

$$\langle w, r(x, j) \rangle - h_S(w) \leq 0, \quad \forall j \in J. \quad (10)$$

Note that the last condition can be written as $\operatorname{val}(w \cdot r) \leq h_S(w)$, where

$$\operatorname{val}(w \cdot r) \triangleq \min_{x \in \Delta(I)} \max_{j \in J} \langle w, r(x, j) \rangle,$$

the minimax value of the game with the scalar payoff that is obtained by projecting the reward vectors $r(i, j)$ onto w . Consequently, a mixed action x that satisfies (10) can be computed as the minimax strategy for the agent in this game.

3.2 The General Algorithm

The proposed algorithm builds on the following idea. First, we employ an OCO algorithm to generate a sequence of *steering vectors* $w_t \in B_2$, so that

$$\sum_{t=1}^T (\langle w_t, r_t \rangle - h_S(w_t)) \geq T \max_{w \in B_2} \{\langle w, \bar{r}_T \rangle - h_S(w)\} - a(T), \quad (11)$$

where $r_t = r(x_t, j_t)$ is considered an arbitrary vector that is revealed after w_t is specified, and $a(T) = o(T)$. Next, given w_t , we choose x_t that satisfies (10), so that $\langle w_t, r_t \rangle - h_S(w_t) \leq 0$. Using this inequality in (11), and observing the distance formula (8), yields

$$d(\bar{r}_T, S) \leq \frac{a(T)}{T} \rightarrow 0.$$

To secure (11), observe that the function $f(w; r) = -\langle w, r \rangle + h_S(w)$ is convex in w for each vector r . Therefore, an OCO algorithm can be applied to the sequence of convex functions $f_t(w) = -\langle w, r_t \rangle + h_S(w)$, where $r_t = r(x_t, j_t)$ is considered an arbitrary vector which is revealed only after w_t is specified. Applying an OCO algorithm \mathcal{A} with $\text{Regret}_T(\mathcal{A}) \leq a(T)$ to this setup, we obtain a sequence (w_t) such that

$$\sum_{t=1}^T f_t(w_t) \leq \min_{w \in B_2} \sum_{t=1}^T f_t(w) + a(T),$$

where

$$\begin{aligned} \sum_{t=1}^T f_t(w_t) &= - \sum_{t=1}^T (\langle w_t, r_t \rangle - h_S(w_t)), \\ \sum_{t=1}^T f_t(w) &= - \sum_{t=1}^T (\langle w, r_t \rangle - h_S(w)) = -T(\langle w, \bar{r}_T \rangle - h_S(w)). \end{aligned}$$

This clearly implies (11).

The discussion above leads to the following approachability meta-algorithm.

Algorithm 1 (Approachability Meta-Algorithm Based on OCO)

Given: A closed, convex and approachable set S ; a procedure (e.g., a linear program) to compute x , for a given vector w , so that (10) is satisfied; an OCO algorithm \mathcal{A} for the functions $f_t(w) = -\langle w, r_t \rangle + h_S(w)$, with $\text{Regret}_T(\mathcal{A}) \leq a(T)$.

Repeat for $t = 1, 2, \dots$:

1. Obtain w_t from the OCO algorithm applied to the convex functions $f_k(w) = -\langle w, r_k \rangle + h_k(w)$, $k \leq t-1$, so that inequality (11) is satisfied.

2. Choose x_t according to (10), so that $\langle w_t, r(x_t, j) \rangle - h_S(w_t) \leq 0$ holds for all $j \in J$.

3. Observe Nature's action j_t , and set $r_t = r(x_t, j_t)$.

Proposition 5 *For the algorithm above,*

$$d(\bar{r}_T, S) \leq \frac{a(T)}{T}$$

is satisfied for all $T \geq 1$ and any sequence (j_1, j_2, \dots) of Nature's actions.

Proof: As observed above, application of the OCO algorithm implies (11), so that

$$\begin{aligned} d(\bar{r}_T, S) &= \max_{w \in B_2} \{ \langle w, \bar{r}_T \rangle - h_S(w) \} \\ &\leq \frac{1}{T} \sum_{t=1}^T (\langle w_t, r_t \rangle - h_S(w_t)) + \frac{a(T)}{T} \leq \frac{a(T)}{T}. \end{aligned}$$

□

To recap, any OCO algorithm that guarantees (11) with $\frac{a(T)}{T} \rightarrow 0$, induces an approachability strategy with rate of convergence $\frac{a(T)}{T}$.

Remark 1 (Convex Cones) *The approachability algorithm developed in [1] starts with a target sets S that are restricted to be convex cones. For S a closed convex cone, the support function is given by*

$$h_S(w) = \begin{cases} 0 & : w \in S^o \\ \infty & : w \notin S^o \end{cases}$$

where S^o is the polar cone of S . The required inequality in (11) therefore reduces to

$$\sum_{t=1}^T \langle w_t, r_t \rangle \geq T \max_{w \in B_2 \cap S^o} \langle w, \bar{r}_T \rangle - a(T).$$

The sequence (w_t) can be obtained in this case by applying an online linear optimization algorithm restricted to $w_t \in B_2 \cap S^o$. This is the algorithm proposed in [1].

The extension to general convex sets is handled there by lifting the problem to a $(d+1)$ -dimensional space, with payoff vector $r'(x, y) = (\kappa, r(x, y))$ and target set $S' = \text{cone}(\{\kappa\} \times S)$, where $\kappa = \max_{s \in S} \|s\|$, for which it holds that $d(u, S) \leq 2d(u', S')$. For further details see [1].

3.3 An OGD-based Approachability Algorithm

As a concrete example, let us apply the Online Gradient Descent algorithm specified in (4) to our problem. With $W = B_2$ and $f_t(w) = -(\langle w, r_t \rangle - h_S(w))$, we obtain in step 1 of Algorithm 1,

$$w_{t+1} = \text{Proj}_{B_2}\{w_t + \eta_t(r_t - y_t)\}, \quad y_t \in \partial h_S(w_t).$$

Observe that $\text{Proj}_{B_2}(v) = v / \max\{1, \|v\|\}$, and (e.g., Corollary 8.25 in [14])

$$\partial h_S(w) = \underset{s \in S}{\operatorname{argmax}} \langle s, w \rangle.$$

To evaluate the convergence rate in (5), observe that $\text{diam}(B_2) = 2$, and, since $y_t \in S$, $\|g_t\| = \|r_t - y_t\| \leq \|\mathcal{R} - S\|$, where $\mathcal{R} = \{r(x, y)\}_{x \in \Delta(I), y \in \Delta(J)}$ is the reward set. Assuming for the moment that the goal set S is bounded, we obtain

$$d(\bar{r}_T, S) \leq \frac{b(\eta)}{\sqrt{T}}, \quad \text{with } b(\eta) = \frac{4}{\eta} + 2\eta\|\mathcal{R} - S\|^2.$$

For $\eta = \sqrt{2}/\|\mathcal{R} - S\|$, we thus obtain $b(\eta) = 4\sqrt{2}\|\mathcal{R} - S\|$.

If S is not bounded, it can always be intersected with \mathcal{R} (without affecting its approachability), yielding $\|\mathcal{R} - S\| \leq \text{diam}(\mathcal{R})$. This amounts to modifying the choice of y_t in the algorithm to

$$y_t \in \partial h_{S \cap \mathcal{R}}(w_t) = \underset{y \in S \cap \mathcal{R}}{\operatorname{argmax}} (y, w).$$

Alternatively, one may restrict attention (by projection) to vectors w_t in the set $\{w \in B_2 : h_S(w) < \infty\}$, similarly to the case of convex cones mentioned in Remark 1 above; we will not go into further details here.

4 Blackwell's Algorithm and (R)FTL

We next examine the relation between Blackwell's approachability algorithm and the present OCO-based framework. We first show that Blackwell's algorithm coincides with OCO-based approachability when FTL is used as the OCO algorithm. We use this equivalence to establish fast (logarithmic) convergence rates for Blackwell's algorithm when the target set S has a smooth boundary. Interestingly, this equivalence does not provide a convergence result for general convex sets. To complete the picture, we show that Blackwell's algorithm can more generally be obtained via a *regularized* version of FTL, which leads to an alternative proof of convergence of the algorithm in the general case.

4.1 Blackwell's algorithm as FTL

Recall Blackwell's algorithm as specified in Theorem 1, namely x_{t+1} is chosen as a mixed action that satisfies (2) for $u = u_S(\bar{r}_t)$.

Lemma 6 For $f_t(w) = -\langle w, r_t \rangle + h_S(w)$,

$$\operatorname{argmin}_{w \in B_2} \sum_{k=1}^t f_k(w) = \begin{cases} u_S(\bar{r}_t) & : \bar{r}_t \notin S \\ 0 & : \bar{r}_t \in S \end{cases}.$$

Proof: Observe that $\sum_{k=1}^t f_k(w) = -t(\langle w, \bar{r}_t \rangle - h_S(w))$, so that

$$\operatorname{argmin}_{w \in B_2} \sum_{k=1}^t f_k(w) = \operatorname{argmax}_{w \in B_2} \{\langle w, \bar{r}_t \rangle - h_S(w)\}.$$

The required equality now follows from (9). \square

Comparing to (6), with $R_t \equiv 0$, it may be seen that the sequence of projection directions $u_S(\bar{r}_t)$ in Blackwell's algorithm coincides with the sequence (w_t) that is obtained by applying the FTL algorithm to the functions (f_t) over $w \in B_2$. It follows that Blackwell's algorithm is identical to Algorithm 1 with this choice of the OCO algorithm.

To establish convergence of Blackwell's algorithm via this equivalence, one needs to show that FTL guarantees the regret bound in (11) for an arbitrary reward sequence $(r_t) \subset \mathcal{R}$, with a sublinear rate sequence $a(T)$. It is well known, however, that (unregularized) FTL does not guarantee sublinear regret, without some additional assumptions on the function f_t . A simple counter-example, reformulated to the present case, is devised as follows: Let $S = \{0\} \subset \mathbb{R}$, so that $h_S(w) = 0$, and suppose that $r_1 = -1$ and $r_t = 2(-1)^t$ for $t > 1$. Since $w_t = \operatorname{sign}(\bar{r}_{t-1})$ and $\operatorname{sign}(r_t) = -\operatorname{sign}(\bar{r}_{t-1})$, we obtain that $f_t(w_t) = -r_t w_t = 1$, leading to a linearly-increasing regret.

The failure of FTL in this example is clearly due to the fast changes in the predictors w_t . We now add some smoothness assumptions on the set S that can mitigate such abrupt changes.

Assumption 1 Let S be a compact and convex set. Suppose that the boundary ∂S of S is smooth with curvature bounded by κ_0 , namely:

$$\|\vec{n}(s_1) - \vec{n}(s_2)\| \leq \kappa_0 \|s_1 - s_2\| \quad \text{for all } s_1, s_2 \in \partial S, \quad (12)$$

where $\vec{n}(s)$ is the unique unit outer normal to S at $s \in \partial S$.

For example, for a closed Euclidean ball of radius ρ , (12) is satisfied with equality for $\kappa_0 = \rho^{-1}$. The assumed smoothness property may in fact be formulated in terms of an interior sphere

condition: For any point in $s \in S$ there exists a ball $B(\rho) \subset S$ with radius $\rho = \kappa_0^{-1}$ such that $s \in B(\rho)$.

Proposition 7 *Let Assumption 1 hold. Consider Blackwell's algorithm as specified in Theorem 1, and denote $w_t = u_S(\bar{r}_{t-1})$ (with w_1 arbitrary). Then, for any time $T \geq 1$ such that $\bar{r}_T \notin S$, (11) holds with*

$$a(T) = C_0(1 + \ln T), \quad (13)$$

where $C_0 = \text{diam}(\mathcal{R}) \|\mathcal{R} - S\| \kappa_0$, $C_1 = \|\mathcal{R} - S\|$, and $\ln(\cdot)$ is the natural logarithm. Consequently,

$$d(\bar{r}_T, S) \leq C_0 \frac{1 + \ln T}{T}, \quad T \geq 1. \quad (14)$$

Proof: See the Appendix.

The last result establishes a fast convergence rate (of order $\log T/T$) for Blackwell's approachability algorithm, under the assumed smoothness of the target set. We observe that in the stochastic version of the algorithm, which is based on the rewards $r(i_t, j_t)$ rather than $r(x_t, j_t)$, the convergence is still of order $T^{-1/2}$ due to the added stochastic effect (unless all mixed actions x_t happen to be pure). We also note that logarithmic convergence rates for OCO algorithms were derived in [10], under strong convexity conditions on the function f_t . Finally, conditions for fast approachability (of order T^{-1}) were derived in [13], but are of different nature than the above.

4.2 Blackwell's algorithm as RFTL

The smoothness requirement in Assumption 1 precludes such important target sets as polyhedra and cones. As observed above, in absence of such additional smoothness properties the interpretation of Blackwell's algorithm through an FTL scheme does not imply its convergence, as the regret of FTL (and the corresponding bound $a(T)$ in (11)) might increase linearly in general.

To address the general case, we show next that the Blackwell's algorithm can be identified more generally with a *regularized* version of FTL. This algorithm does guarantee an $O(\sqrt{T})$ regret in (11), and consequently leads to the standard $O(T^{-1/2})$ rate of convergence of Blackwell's approachability algorithm.

Our starting point is the following observation:

Lemma 8 *For $f_k(w) = -\langle w, r_k \rangle + h_S(w)$, $1 \leq k \leq t$, and any $\rho_t > 0$,*

$$w_{t+1} \triangleq \underset{w \in B_2}{\operatorname{argmin}} \left\{ \sum_{k=1}^t f_k(w) + \frac{\rho_t}{2} \|w\|^2 \right\} = \begin{cases} \beta_t u_S(\bar{r}_t) & : \bar{r}_t \notin S \\ 0 & : \bar{r}_t \in S \end{cases}. \quad (15)$$

where $\beta_t = \min\{1, \frac{t}{\rho_t} d(\bar{r}_t, S)\} > 0$.

Proof: Recall that $\sum_{k=1}^t f_k(w) = -t(\langle w, \bar{r}_t \rangle - h_S(w))$, so that

$$\operatorname{argmin}_{w \in B_2} \left\{ \sum_{k=1}^t f_k(w) + \frac{\rho_t}{2} \|w\|^2 \right\} = \operatorname{argmax}_{w \in B_2} \left\{ \langle w, \bar{r}_t \rangle - h_S(w) - \frac{\rho_t}{2t} \|w\|^2 \right\}.$$

To compute the right-hand side, we first maximize over $\{w : \|w\| = \beta\}$, and then optimize over $\beta \in [0, 1]$. Denote $r = \bar{r}_t$, and $\eta = \rho_t/t$. Similarly to Lemma 6,

$$\operatorname{argmax}_{\|w\|=\beta} \left\{ \langle w, r \rangle - h_S(w) - \frac{\eta}{2} \|w\|^2 \right\} = \operatorname{argmax}_{\|w\|=\beta} \left\{ \langle w, r \rangle - h_S(w) \right\} = \begin{cases} \beta u_S(r) & : r \notin S \\ 0 & : r \in S \end{cases}.$$

Now, for $r \notin S$,

$$\max_{\|w\|=\beta} \left\{ \langle w, r \rangle - h_S(w) - \frac{\eta}{2} \|w\|^2 \right\} = \beta d(r, S) - \frac{\eta}{2} \beta^2.$$

Maximizing the latter over $0 \leq \beta \leq 1$ gives $\beta^* = \min\{1, \frac{d(r, S)}{\eta}\}$. Substituting back r and η gives (15). \square

Equation (15) defines an RTFL algorithm with quadratic regularization. When used for the OCO part in Algorithm 1, the resulting scheme turns out to be equivalent to Blackwell's algorithm. Indeed, the minimum in (15) is attained by the same unit vector $u_S(\bar{r}_t)$ that appears in Theorem 1, scaled by a positive constant. That scaling does not affect the choice of x_t according to (10), as the support function $h_S(w)$ is positive homogeneous. However, this scaling does induce sublinear-regret for the OLO algorithm, and consequently convergence of the approachability algorithm. This is summarized as follows.

Proposition 9 *Let S be a convex and compact set. Consider the RTFL algorithm specified in equation (15), with $\rho_t = \rho\sqrt{t}$, $\rho > 0$. The regret of this algorithm is bounded by*

$$\operatorname{Regret}_T(\text{RTFL}) \leq \left(\frac{2L_f^2}{\rho} + \rho \right) \sqrt{T} + \frac{2L_f^2}{\rho} + L_f \ln(4T - 3) \triangleq a_0(T),$$

where $L_f = \|\mathcal{R} - S\|$. Consequently, if this RTFL algorithm is used in step 1 of Algorithm 1 to provide w_t , we obtain

$$d(\bar{r}_T, S) \leq \frac{a_0(T)}{T} = O(T^{-\frac{1}{2}}), \quad T \geq 1. \quad (16)$$

Proof: The regret bound follows from the one in Proposition 3, evaluated for $f_t(w) = -\langle r_t, w \rangle + h_S(w)$, $W = B_2$, $R(w) = \|w\|^2$, and $\rho_t = \rho_0\sqrt{t}$. Recalling that $\partial f_t(w) = -r_t + \operatorname{argmax}_{s \in S} \langle w, s \rangle$, the Lipschitz constant of f_t is upper bounded by $\|\mathcal{R} - S\| \triangleq L_f$. Furthermore, $R_{\max} = 1$ and $L_R = 2$. Therefore,

$$\operatorname{Regret}_T(\text{RTFL}) \leq 2L_f \sum_{t=1}^T \frac{L_f + 2\rho(\sqrt{t} - \sqrt{t-1})}{\rho(\sqrt{t} + \sqrt{t-1})} + \rho\sqrt{T}.$$

Upper bounding the sums with corresponding integrals gives the stated regret bound. The second part now follows directly from Proposition 5. \square

With $\rho = \sqrt{2}L_f$, we obtain in (16) the convergence rate

$$d(\bar{r}_T, S) \leq \frac{2\sqrt{2}\|\mathcal{R} - S\|}{\sqrt{T}} + o\left(\frac{1}{\sqrt{T}}\right).$$

We emphasize that the algorithm discussed in this section is equivalent to Blackwell’s algorithm, hence its convergence is well known. The proof of convergence here is certainly not the simplest, nor does it lead to the best constants in the convergence rate. Indeed, Blackwell’s proof (which recursively bounds the square distance $d(\bar{r}_T, S)^2$) leads to the bound $d(\bar{r}_T, S) \leq \frac{\|\mathcal{R} - S\|}{\sqrt{T}}$. Rather, our main purpose here was to provide an alternative view and analysis of Blackwell’s algorithm, which rely on a standard OCO algorithm. Nonetheless, the logarithmic convergence rate that was obtained under the smoothness Assumption 1 appears to be new.

Acknowledgements

The author wishes to thank Elad Hazan for helpful comments on a preliminary version of this work. This research was supported by the Israel Science Foundation grant No. 1319/11.

References

- [1] J. Abernethy, P. L. Bartlett, and E. Hazan. Blackwell approachability and low-regret learning are equivalent. In *Proceedings of the 24th Conference on Learning Theory (COLT’11)*, pages 27–46, Budapest, Hungary, June 2011.
- [2] A. Bernstein and N. Shimkin. Response-based approachability with applications to generalized no-regret problems. To appear in *Journal of Machine Learning Research*, 2015.
- [3] D. Blackwell. Controlled random walks. In *Proceedings of the International Congress of Mathematicians*, volume III, pages 335–338, 1954.
- [4] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.
- [5] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.

- [6] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, New York, NY, 2006.
- [7] J. Hannan. Approximation to Bayes risk in repeated play. *Contributions to the Theory of Games*, 3:97–139, 1957.
- [8] S. Hart and A. Mas-Colell. A general class of adaptive strategies. *Journal of Economic Theory*, 98:26–54, 2001.
- [9] E. Hazan. The convex optimization approach to regret minimization. In S. Sra et al., editor, *Optimization for Machine Learning*, chapter 10. MIT Press, Cambridge, MA, 2012.
- [10] E. Hazan, A. Agarwal, and S. Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.
- [11] M. Maschler, E. Solan, and S. Zamir. *Game Theory*. Cambridge University Press, Cambridge, UK, 2013.
- [12] V. Perchet. Approachability, regret and calibration: Implications and equivalences. *Journal of Dynamics and Games*, 1:181–254, 2014.
- [13] V. Perchet and S. Mannor. Approachability, fast and slow. In *Proc. COLT 2013: JMLR Workshop and Conference Proceedings*, volume 30, pages 474–488, 2013.
- [14] R.T. Rockafellar and R. Wets. *Variational Analysis*. Springer-Verlag, 1997.
- [15] S. Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4:107–194, 2011.
- [16] M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning (ICML '03)*, pages 928–936, 2003.

Appendix

Proof of Proposition 3: We follow the outline of the proof of Lemma 2.10 in [15], modified to accommodate a non-constant regularization sequence ρ_t . The starting point is the inequality, proved by induction,

$$\sum_{t=1}^T (f_t(w_t) - f_t(u)) \leq \sum_{t=1}^T (f_t(w_t) - f_t(w_{t+1})) + \rho_t R(u), \quad (17)$$

which holds for any $u \in W$. Therefore,

$$\sum_{t=1}^T (f_t(w_t) - f_t(u)) \leq L_f \sum_{t=1}^T \|w_t - w_{t+1}\| + \rho_t R(u). \quad (18)$$

Denote $F_t(w) = \sum_{k=1}^{t-1} f_k(w) + \rho_{t-1} R(w)$. Then F_t is ρ_{t-1} -strongly convex, and w_t is its maximizer by definition. Hence, it holds generally that

$$F_t(u) \geq F_t(w_t) + \frac{\rho_{t-1}}{2} \|u - w_t\|^2,$$

and in particular,

$$F_t(w_{t+1}) \geq F_t(w_t) + \frac{\rho_{t-1}}{2} \|w_{t+1} - w_t\|^2, \quad (19)$$

$$F_{t+1}(w_t) \geq F_{t+1}(w_{t+1}) + \frac{\rho_t}{2} \|w_t - w_{t+1}\|^2. \quad (20)$$

Summing and cancelling terms, we obtain

$$f_t(w_t) - f_t(w_{t+1}) + (\rho_t - \rho_{t-1})(R(w_t) - R(w_{t+1})) \geq \frac{\rho_t + \rho_{t-1}}{2} \|w_{t+1} - w_t\|^2.$$

But the left-hand side is upper-bounded by $(L_f + (\rho_t - \rho_{t-1})L_R)\|w_{t+1} - w_t\|$, which implies that

$$\|w_{t+1} - w_t\| \leq 2 \frac{L_f + (\rho_t - \rho_{t-1})L_R}{\rho_t + \rho_{t-1}}.$$

Substituting in (18) gives the bound stated in the Proposition. \square

Proof of Proposition 7: We first observe that the regret bound in (13) implies (14). Indeed, for $\bar{r}_T \notin S$, $d(\bar{r}_T, S) \leq a(T)/T$ follows as in Proposition 5, while if $\bar{r}_T \in S$ then $d(\bar{r}_T, S) = 0$ and (14) holds trivially.

We proceed to establish the logarithmic regret bound in (13). Let $f_t(w) = -\langle w, r_t \rangle + h_S(w)$, $W = B_2$, and denote

$$\text{Regret}_T(f_{1:T}) = \sum_{t=1}^T f_t(w_t) - \min_{w \in W} \sum_{t=1}^T f_t(w) = \sum_{t=1}^T (f_t(w_t) - f_t(w_{T+1})). \quad (21)$$

A standard induction argument (e.g., Lemma 2.1 in [15]) verifies that

$$\sum_{t=1}^T (f_t(w_t) - f_t(u)) \leq \sum_{t=1}^T (f_t(w_t) - f_t(w_{t+1})) \quad (22)$$

holds for any $u \in W$, and in particular for $u = w_{T+1}$. It remains to upper-bound the differences in the last sum.

Consider first the case where $\bar{r}_t \notin S$ for all $1 \leq t \leq T$. We first show that $\|w_t - w_{t+1}\|$ is small, which implies the same for $|f_t(w_t) - f_t(w_{t+1})|$. By its definition, $w_{t+1} = u_S(\bar{r}_t)$, the unit vector

pointing to \bar{r}_t from $c_t \triangleq \text{Proj}_S(\bar{r}_t)$, which clearly coincides with the outer unit normal $\vec{n}(c_t)$ to S at c_t . It follows that

$$\|w_t - w_{t+1}\| = \|\vec{n}(c_{t-1}) - \vec{n}(c_t)\| \leq \kappa_0 \|c_{t-1} - c_t\| \leq \kappa_0 \|\bar{r}_{t-1} - \bar{r}_t\|,$$

where the first inequality follows by Assumption 1, and the second due to the shrinking property of the projection. Substituting $\bar{r}_t = \bar{r}_{t-1} + \frac{1}{t}(r_t - \bar{r}_{t-1})$ obtains

$$\|w_t - w_{t+1}\| \leq \frac{\kappa_0}{t} \|r_t - \bar{r}_{t-1}\| \leq \frac{\kappa_0}{t} \text{diam}(\mathcal{R}). \quad (23)$$

Next, observe that for any pair of unit vectors w_1 and w_2 ,

$$\begin{aligned} f_t(w_1) - f_t(w_2) &= -\langle w_1 - w_2, r_t \rangle + h_S(w_1) - h_S(w_2) \\ &= -\langle w_1 - w_2, r_t \rangle + \max_{s \in S} \langle w_1, s \rangle - \max_{s \in S} \langle w_2, s \rangle \\ &\leq -\langle w_1 - w_2, r_t \rangle + \langle w_1, s_1 \rangle - \langle w_2, s_1 \rangle \\ &= \langle w_1 - w_2, s_1 - r_t \rangle \leq \|w_1 - w_2\| \|\mathcal{R} - S\|, \end{aligned}$$

where $s_1 \in S$ attains the first maximum. Since the same bound holds for $f_t(w_2) - f_t(w_1)$, it holds also for the absolute value. In particular,

$$|f_t(w_t) - f_t(w_{t+1})| \leq \|w_t - w_{t+1}\| \|\mathcal{R} - S\|, \quad (24)$$

and together with (23) we obtain

$$|f_t(w_t) - f_t(w_{t+1})| \leq \frac{\kappa_0}{t} \text{diam}(\mathcal{R}) \|\mathcal{R} - S\| = \frac{C_0}{t}.$$

Substituting in (22) and summing over t^{-1} yields the regret bound

$$\text{Regret}_T(f_{1:T}) \leq C_0(1 + \ln T). \quad (25)$$

We next extend this bound to case where $\bar{r}_t \in S$ for some t . In that case $w_{t+1} = 0$, and $w_t - w_{t+1}$ may not be small. However, since $f_t(0) = 0$, such terms will not affect the sum in (22). Recall that we need to establish (13) for T such that $br_T \notin S$. In that case, any time t for which $\bar{r}_t \in S$ is followed by some time $m \leq T$ with $\bar{r}_m \notin S$. Let $1 \leq k < m \leq T$ be indices such that $\bar{r}_k, \dots, \bar{r}_{m-1} \in S$, but $\bar{r}_{k-1} \notin S$ (or $k = 1$) and $\bar{r}_m \notin S$. Then $w_{k+1}, \dots, w_m = 0$, and

$$\sum_{t=k}^m (f_t(w_t) - f_t(w_{t+1})) = f_k(w_k) - f_m(w_{m+1}).$$

Proceeding as above, we obtain similarly to (23),

$$\|w_k - w_{m+1}\| \leq \kappa_0 \|\bar{r}_{k-1} - \bar{r}_m\| \leq \text{diam}(\mathcal{R}) \sum_{t=k}^{m-1} \frac{\kappa_0}{t},$$

and the regret bound in (25) may be obtained as above. \square